A unified model of spatial and object attention based on inter-cortical biased competition [1]

Gustavo Deco* & Tai Sing Lee†

* *Siemens AG, Corporate Technology,*
ZTIK 4 Otto-Hahn-Ring 6, 81739 Munich, Germany
†*Department of Computer Science & Center for the Neural Basis of Cognition*
Carnegie Mellon University Pittsburgh, PA 15213, U.S.A.

Abstract

We present a physiologically constrained neural dynamical model of the visual system for the organization of attention and its mediation of object recognition and visual search. In this model, spatial and feature attention are mediated by a single neural mechanism involving the interaction of the ventral and the dorsal streams with the early visual cortex. The model consists of three representative modules which encode object classes, spatial locations, and elementary features respectively. These modules are coupled together in a neural dynamical system in the framework of biased competition. The system can be made to operate in either a spatial or an object attention mode by introducing a top-down bias to either the dorsal or the ventral stream modules. In this system, translation invariant object recognition and object spatial localization arise from the interaction among the modules, with the early visual areas playing a key role in mediating such interaction.

Key words: attention; visual search; V1; extrastriate cortex; ; neural model.

1 Introduction

Visual attention can function in two distinct modes: spatial focal attention that can be visualized as a spotlight that 'illuminate' a certain location of visual space for focused visual analysis [4]; spatially distributed object attention with which a target object can be searched in parallel over a large visual space [7]. Duncan [2] proposed that the two modes of operation are both manifestation of a top-down selection process. In spatial attention, the selection

[1] This research was supported by NSF LIS 9720350, NSF CAREER 9984706 NIH Vision Research core grant EY08098, and a grant from Siemens, AG. Corresponding author: T.S. Lee tai@cnbc.cmu.edu

is focal in the spatial dimension and diffuse in feature dimension; while in object attention, the selection is focal in the feature dimension and diffuse in the spatial dimension. In recent years, a number of neurophysiological studies [5,10,11,13] have provided insights to the neural basis of spatial attention and object attention. Several computational models have been advanced to account for aspects of either spatial attention or object attention [12,14,15,17]. Yet, none of these models, except [1], attempt to integrate spatial and object attention mechanisms into a unitary system.

The system described in this paper is built upon earlier models on biased-competition in the ventral stream [2, 13, 15], and on interactive processes in visual processing [3,9]. Our main proposal here is that translation invariant object recognition and visual search can be accomplished through the interaction between the interaction of the ventral-stream module and the dorsal-stream module via the early visual cortex such as V1 and V2. Early visual cortex plays a special role in mediating and integrating information fed back from the various expert modules in the extrastriate cortex, as proposed in the high-resolution buffer theory of V1 [6]. In this framework, spatial attention and object attention mechanisms can be integrated in an unified system.

2 Methods

A neural dynamical system with three interacting modules has been implemented. The three cortical modules are the early visual cortical module (V1/V2), a ventral-stream module (V4/IT), and a dorsal-stream module (PP/PO). Spatial or object attention are generated by top-down bias input to the dorsal-stream module or the ventral-stream module respectively, and the two streams interact through the bidirectional connection between these modules and the early visual cortex, as shown in Figure 1.

The unit in each module represents a pool of neurons with similar properties. Its activity is described using mean field approximation [16]. In this formulation, each unit i is characterized by two variables: its activation x_i (average firing rate of the pool) and an input current I_i, characteristic for all cells in pool i, satisfying the following input-output relationship:

$$x_i(t) = F(I_i(t)) = \frac{1}{T_r - \tau \log(1 - \frac{1}{\tau I_i(t)})}$$

where T_r is the cell's absolute refractory period (e.g. 1 msec) and τ is the membrane time constant. Let there be m neuronal pools in each cortical module. Each excitatory pool's dynamics is then described by, for $i = 1, ...m$,

$$\tau \frac{\partial}{\partial t} I_i(t) = -I_i + aF(I_i(t)) - bF(I^I(t)) + I_i^B(t) + I_i^T(t) + I_o + \nu$$

The first term is a habituation decay term. The second term represents the recurrent self excitation for maintaining the activity of the pool. It mediates

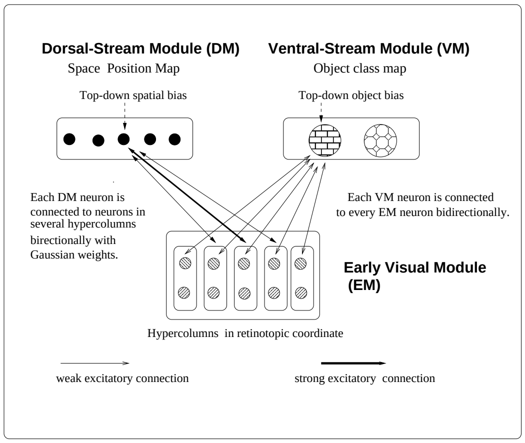

Dorsal-Stream Module (DM)
Space Position Map
Ventral-Stream Module (VM)
Object class map
Top-down spatial bias
Top-down object bias
Each DM neuron is connected to neurons in several hypercolumns birectionally with Gaussian weights.
Each VM neuron is connected to every EM neuron bidirectionally.
Early Visual Module (EM)
Hypercolumns in retinotopic coordinate
weak excitatory connection
strong excitatory connection

Fig. 1. (a) A schematic diagram of the model. The model contains three modules: the early visual module (EM), the ventral stream module (VM) and the dorsal stream module (DM). The early visual module contains orientation-selective complex cells and hypercolumns as in the primary visual cortex. The ventral stream module contains neuronal pools encoding specific object classes as in the inferotemporal cortex. The dorsal stream module contains a map encoding positions in the retinotopic coordinate. The early module and the ventral module are connected with symmetrical connections developed with Hebbian learning. The early module and the dorsal module are connected with symmetrically localized connections modeled with Gaussian weights. Competitive interaction within each module is mediated by inhibitory pools. Connection between modules are excitatory, biasing the competitive dynamics in each module. Concentration of neural activities to an individual pool in the ventral module corresponds to object recognition. Concentration of neural activities to a unit in a dorsal module corresponds to object localization. The early module provides a buffer for the ventral and the dorsal modules to interact.

the cooperative interaction among neurons within each unit. The third term is the inhibitory input from the inhibitory neuronal pool. I_i^B is the specific bottom-up input to pool i from a lower cortical module, and I_i^T is the specific top-down bias input from higher cortical modules. I_o and ν are diffuse spontaneous background input and an additive Gaussian noise to the system respectively.

The inhibitory neuronal pool integrates information from all the excitatory pools within each module and feeds back unspecific inhibition to each of the excitatory pools. It mediates competitive normalizing interaction among the neuronal pools within the module. Its dynamics is given by,

$$\tau \frac{\partial}{\partial t} I^I(t) = -I^I - dF(I^I(t)) + \sum_{i=1}^{m} F(I_i(t))$$

These two dynamical equations are the fundamental components in our system.

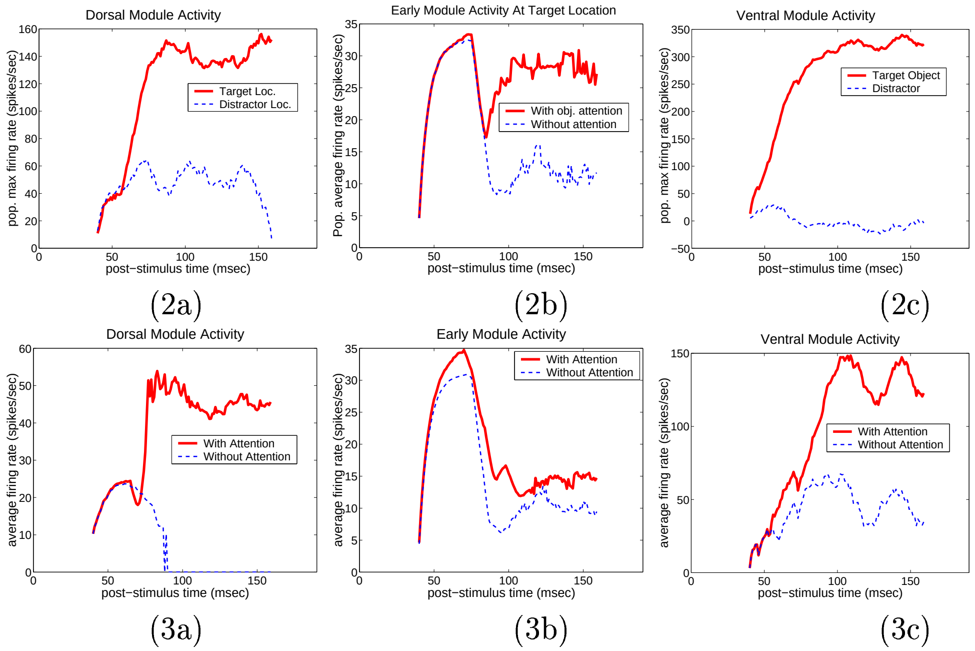
The early visual module (EM), which models areas V1 and V2, contains 33 x 33 hypercolumns, covering a 66 x 66 pixel scene. Each of the hypercolumn contains 24 pools, 8 orientations and 3 scales, of feature detectors, modeled by power modulus of Gabor wavelet responses to the input images. For simplicity, we allow global normalization within units at each scale within the early visual module. Hence, there are 26136 excitatory pools and 3 inhibitory pools in the early visual module.

The dorsal stream module (DM), a lattice of 66 x 66 units, computes and maintains a spatial map of an object's location. It helps to direct the spotlight of spatial attention on the early visual module. Neurophysiologically, it might include a number of cortical areas in the dorsal stream such as V3a, LIP and PO. This module receives top-down bias input from the prefrontal cortex which specifies the locus of spatial attention. Each node on DM lattice is represented a pool of neurons that are reciprocally connected to a set of hypercolumns in V1 with a Gaussian spatial spread. The activation of each DM node indicates spatial attention allocation to the hypercolumns under its feedback influence. There is one common inhibitory pool that mediates competitive interactions among these 4356 neuronal pools.

The ventral stream module (VM) represents memories of object classes and is responsible for object categorization and recognition. Neurophysiologically, this module might include TEO, TE and other areas of the inferotemporal cortex. The module in our model has two excitatory pools of neurons, each representing a particular object. A top-down bias input from prefrontal cortex to the pool will select a particular object, initiating the effect of object attention in a visual search task. Each VM unit is fully and symmetrically connected to all neuronal pools in the early visual module. Their connection weights are trained by Hebbian learning in a learning phase.

3 Results

The system are tested in two scenarios: object attention and spatial attention. In the object attention mode, as when the animal is looking for a particular object in a visual search task, a top-down bias is imposed onto a VM unit, tilting the balance of competition in VM. As the competition in VM proceeds, the activity in VM will also propagate back to the early visual module through recurrent feedback connection. The hypercolumns in EM that exhibit a response pattern closest to the patterns corresponding to the 'preferred' unit in VM will be facilitated by the feedback and in turn inhibit other hypercolumns in EM through lateral inhibition. Activities in EM will feed forward to the units in DM, where the competitive interaction will further intensify the spatial localization through the competition within DM. The three modules mutually constrain each other's computation, resulting in an increase in the activity of the DM unit at the target position relative to that of the units in the distractor positions (Fig. 2a), and an increase in the activity of the VM unit corresponding to the target relative to the activity of the distractor unit (Fig 2c). Fig 2b shows the activities of the same hypercolumns with and without object attention, indicating the effect of object attention emerges at the late response of the EM neurons.



(2a) (2b) (2c)

(3a) (3b) (3c)

Fig. 2.

In the spatial attention mode, the top down bias is imposed on a unit in the dorsal module's spatial map to specify the locus of spatial attention. This unit will enhance the activities at a specified region in the early visual area. The simulation shows the effect of spatial attentional bias on the DM unit (Fig 3a) and its impact on the early module units (Fig 3b). The effect of spatial attention also emerges at the late response of the neurons. The highlighted hypercolumns provides a stronger bias to the VM unit coding for the object analyzed by these hypercolumns under the attentional spotlight. Figure 3c shows the increase in activity of the VM unit corresponding to a particular object when it is under the spotlight relative to when it is not. The bias from EM will eventually allow the target unit in VM become the winner, corresponding to the recognition of the object by the visual system. By successively introducing top-down bias to different spatial units in the dorsal module, the system can effectively highlight and gate information from different retinotopic locations of the early visual areas to the object recognition module. This mechanism implements translation invariant object recognition without the need for dynamic synaptic modifications in the other models [12,13] such as the shifter circuit or the need for attention gain field cells [14]. The DM map in our model encodes spatial location but is not an explicit saliency map [8,17]. Representation of saliency is distributed across multiple modules in our model.

Our simulation shows that a small enhancement due to object attention or spatial attention at the level of V1 and V2 is sufficient to communicate the bias between VM and DM, causing dramatic symmetry breaking in the higher modules. The mutually constrained computations in the three modules lead to simultaneous localization of the target in the DM map, the identification of target in the VM map, highlighted features in the EM map, binding an object's identify, spatial location and its detailed features into one unified percept in our brain.

References

[1] G.T. Buracas, T.D. Albright & T.J. Sejnowski Varieties of attention: a model of visual search, *Proceeding of 3rd Joint symposium on neural computation*, Institute Neural Computation, 6, (1996), 11-25.

[2] J. Duncan, J, The locus of interference in the perception of simultaneous stimuli, *Psychological Review, 87*, (1980) 272-300.

[3] Grossberg, S. Competitive learning: from interactive activation to adapative resonance. *Cogn Sci*, 11 (1987), 23-63.

[4] H.V. Helmholtz, H.V. *Handbuch der physiologischen Optik.* Leipzig: Voss, 1867.

[5] M. Ito, & C. Gilbert, Attention modulates contextual influences in the primary visual cortex of alert monkeys. *Neuron, 22*, (1999) 593-604.

[6] T.S. Lee, D. Mumford, R. Romero & V.A.F. Lamme, The role of primary visual cortex in higher level vision. *Vision Research* **38**, (1998) 2429-2454.

[7] W. James, *The principles of psychology.* New York: Henry Holt, 1890.

[8] C. Koch & S. Ullman, Shifts in selective visual attention: towards the underlying neural circuitry. In: Vaina, L.M. *Matter of Intelligence*, D Reidel Publishing Company, (1987) 115-141.

[9] J.L. McCelland, & D.E. Rumelhart, An interactive activation model of context effects in letter perception. Part I: an account of basic findings. *Psych. Rev, 88*, (1981) 375-407.

[10] J. Moran, & R. Desimone, Selective attention gates visual processing in the extrastriate cortex. *Science, 229*, (1985) 782-784.

[11] B. Motter, Focal attention produces spatially selective processing in visual cortical areas V1, V2 and V4 in the presence of competing stimuli. *Journal of Neurophysiology* 70 (1993), 909-919.

[12] B. Olshausen, C. Andersen, & D. Van Essen, A neural model for visual attention and invariant pattern recognition. *J. Neuroscience, 13(11)*, (1993) 4700-4719.

[13] J. Reynolds, L. Chelazzi and R. Desimone, Competitive mechanisms sub-serve attention in macaque areas V2 and V4. *Journal of Neuroscience*, 19 (1999) 1736-1753.

[14] E. Salinas, & L. Abbott, Invariant visual perception from attentional gain fields. *J. Neurosphysiology*, 77 (1997), 3267-3272.

[15] M. Usher, & E. Niebur, Modelling the temporal dynamics of IT neurons in visual search: A mechanism for top-down selective attention. *J. Cognitive Neuroscience, 8*, (1996) 311-327.

[16] H. Wilson, & J. Cowan, (1972). Excitatory and inhibitory interaction in localized population of model neurons. *Biological Cybernetics, 12* (1972) 1-24.

[17] J.M. Wolfe, Guided search 2. *Psychonomic Bulletin & Review,* 1 (1994), 202-238.